

# MARKOV KERNELS AND THE CONDITIONAL EXTREME VALUE MODEL

SIDNEY I. RESNICK AND DAVID ZEBER

**ABSTRACT.** The classical approach to multivariate extreme value modelling assumes that the joint distribution belongs to a multivariate domain of attraction. This requires each marginal distribution be individually attracted to a univariate extreme value distribution. An apparently more flexible extremal model for multivariate data was proposed by Heffernan and Tawn under which not all the components are required to belong to an extremal domain of attraction but assumes instead the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components is extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation of the probability of certain risk regions. The original focus on conditional distributions had technical drawbacks but is natural in several contexts. We place this approach in the context of the more general approach using convergence of measures and multivariate regular variation on cones.

## 1. OVERVIEW

The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be phrased conveniently in terms of regular variation of the joint distribution on an appropriate cone; see Das and Resnick [5, Proposition 4.1].

A more flexible model for data realizations of a random vector was proposed by Heffernan and Tawn [11], under which not all the components are required to belong to an extremal domain of attraction. Such a model accomodates varying degrees of asymptotic dependence between pairs of components. Instead of starting from the joint distribution, Heffernan and Tawn assumed the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components was extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation for the probabilities of certain multivariate risk sets. However, focusing on conditional distributions creates problems when taking limits owing to ambiguity regarding the choice of version. So the Heffernan/Tawn approach was reformulated as the *conditional extreme value model* (CEVM) in [5, 6, 10] using regular variation of the joint distributions on a smaller cone than the one employed in multivariate extreme value theory, an approach related to hidden regular variation [12, 13, 15, 17].

Conditional distributions are natural objects in many circumstances, for example if densities exist or if one variable is an explicit function of others. So we return to the Heffernan and Tawn [11] formulation, placing it in a formal context that uses the idea of transition kernels in a domain of attraction developed in [18, 19]. We see how reliance on transition kernels fits with general theory expressed in terms of vague convergence of measures and to what extent the reliance on kernels restricts the class of limit measures.

In order to better fit in with the study of extremes of a random vector, we extend the kernel domain of attraction condition used in [18] beyond standardized regular variation to accomodate

---

*Date:* October 12, 2012.

S. I. Resnick and D. Zeber were partially supported by ARO Contract W911NF-10-1-0289 and NSA Grant H98230-11-1-0193 at Cornell University.

general linear normalization in both the initial state and the distribution of the next state. We examine conditions under which this extends to a CEVM, when combined with a marginal domain of attraction assumption, and we derive explicit formulas for the CEV limit measure in different cases. Also, through a number of revealing examples, we explore the properties of the normalization functions, and technicalities surrounding the choice of version of the conditional distribution and the limit distribution  $G$ .

Section 2 summarizes necessary background, definitions and basic results including when a random vector  $(X, Y)$  satisfies the *conditional extreme value model* (CEVM). Section 3 treats the *standard* case where both  $X$  and  $Y$  can be scaled by the same function and this restriction is weakened in Sections 4 and 5.

We denote by  $\mathbb{M}_+(\mathbb{E})$  the space of Radon measures on a nice space  $\mathbb{E}$  topologized by vague convergence which is written as  $\xrightarrow{v}$  [16]. Weak convergence [2] of probability measures is denoted  $\Rightarrow$ . We write  $\xi \sim G$  to mean a random variable  $\xi$  has distribution  $G(\cdot)$  and, if no confusion can arise, we often use  $G$  to also mean the distribution function  $P[\xi \leq x] = G(x)$ . Alternatively for a random variable  $Y$ , we write  $F_Y$  for the distribution of  $Y$ . The class of regularly varying functions with index  $\rho$  on  $(0, \infty)$  is  $RV_\rho$  [3, 7, 8, 16]. The probability measure degenerate at  $c \in \mathbb{R}$  is  $\epsilon_c(\cdot)$ .

## 2. BACKGROUND

First, we review the basics of extended regular variation, which features prominently in the formulation of the CEVM, as well as concepts of univariate extreme value theory. We then define the conditional extreme value model and discuss its basic properties.

**2.1. Extended Regular Variation.** Regular variation and extended regular variation is important in the mathematical description of extreme and conditional extreme value theory [3, 7, 15, 16, 21]. The pair of functions  $a : (0, \infty) \mapsto (0, \infty)$  and  $f : (0, \infty) \mapsto \mathbb{R}$  are *extended regularly varying* (ERV) with parameters  $\rho, k \in \mathbb{R}$  if as  $t \rightarrow \infty$ ,

$$(2.1) \quad \frac{a(tx)}{a(t)} \longrightarrow x^\rho \quad \text{and} \quad \frac{f(tx) - f(t)}{a(t)} \longrightarrow \psi(x), \quad x > 0,$$

[7, Appendix B.2], where

$$(2.2) \quad \psi(x) = \begin{cases} k\rho^{-1}(x^\rho - 1) & \rho \neq 0 \\ k \log x & \rho = 0 \end{cases}.$$

We will write this as  $a, f \in \text{ERV}_{\rho, k}$  with  $a \in \text{RV}_\rho$ . A useful identity is

$$(2.3) \quad \psi(x^{-1}) = -x^{-\rho}\psi(x).$$

Note that this differs slightly from the usual definition of extended regular variation, which assumes  $k = 1$ . If  $\phi(x) := \lim_{t \rightarrow \infty} (f(tx) - f(t))/a(t)$  exists for  $x > 0$ , then  $a$  is necessarily regularly varying, and  $\phi \equiv \psi$ , the function given in (2.2). Also, the convergences in (2.1) are locally uniform, implying that

$$\frac{a(tx_t)}{a(t)} \longrightarrow x^\rho \quad \text{and} \quad \frac{f(tx_t) - f(t)}{a(t)} \longrightarrow \psi(x) \quad \text{whenever } x_t \rightarrow x > 0.$$

Furthermore, if  $k \neq 0$  we obtain the following properties depending on the value of  $\rho$ . Recall the sign function  $\text{sgn}(u) = u/|u| \mathbf{1}_{\{u \neq 0\}}$ .

- If  $\rho > 0$ , then  $f \cdot \text{sgn}(k) \in \text{RV}_\rho$ , and  $f(t)/a(t) \rightarrow k/\rho$ .
- If  $\rho < 0$ , then  $f(\infty) = \lim_{t \rightarrow \infty} f(t)$  exists finite,  $(f(\infty) - f(t))/a(t) \rightarrow k/|\rho|$  and  $(f(\infty) - f) \cdot \text{sgn}(k) \in \text{RV}_{-|\rho|}$ .

- If  $\rho = 0$ , i.e.,  $a$  is slowly varying, then  $f \in \Pi(a)$ , the  $\Pi$ -varying functions with auxiliary function  $a(\cdot)$  [7, Appendix B.2]). Suppose  $k > 0$ . Then  $f(\infty) \leq \infty$  exists. If  $f(\infty) = \infty$ , then  $f \in \text{RV}_0$  and  $f(t)/a(t) \rightarrow \infty$ . If  $f(\infty) < \infty$ , then  $f(\infty) - f \in \text{RV}_0$ , and  $(f(\infty) - f(t))/a(t) \rightarrow \infty$ . If  $k < 0$ , then  $-f$  has these properties.

**2.2. Domains of Attraction.** For  $\gamma \in \mathbb{R}$ , define  $\mathbb{E}_\gamma = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$  so that

$$(2.4) \quad \mathbb{E}_\gamma = \begin{cases} (-\gamma^{-1}, \infty) & \gamma > 0 \\ (-\infty, \infty) & \gamma = 0 \\ (-\infty, |\gamma|^{-1}) & \gamma < 0 \end{cases}.$$

The distribution  $F_Y$  of a random variable  $Y$  is in the *domain of attraction* of an extreme value distribution  $G_\gamma$  for some  $\gamma \in \mathbb{R}$ , written  $F_Y \in D(G_\gamma)$ , if there exist functions  $a(t) > 0$  and  $b(t) \in \mathbb{R}$  such that

$$F_Y^t(a(t)y + b(t)) \longrightarrow G_\gamma(y)$$

weakly as  $t \rightarrow \infty$ , where  $G_\gamma(y) = \exp\{-(1 + \gamma y)^{-1/\gamma}\}$  for  $y \in \mathbb{E}_\gamma$  [7, 16]. This can be reformulated in terms of the tail of the distribution  $F_Y$  as

$$(2.5) \quad t \mathbb{P}[Y - b(t)/a(t) > y] \longrightarrow (1 + \gamma y)^{-1/\gamma}, \quad y \in \mathbb{E}_\gamma.$$

If  $\gamma = 0$ , we interpret the limit as  $e^{-y}$ .

If (2.5) holds for some functions  $a$  and  $b$ , then it holds for [7, Theorem 1.1.6, p. 10])

$$(2.6) \quad b(t) = \left( \frac{1}{1 - F_Y} \right)^{\leftarrow}(t) = F_Y^{\leftarrow}(1 - t^{-1}),$$

where  $g^{\leftarrow}$  is the left-continuous inverse of the nondecreasing function  $g$ . By inversion, (2.5) yields

$$(2.7) \quad \frac{b(tx) - b(t)}{a(t)} \rightarrow \frac{x^\gamma - 1}{\gamma} \mathbf{1}_{\{\gamma \neq 0\}} + \log x \mathbf{1}_{\{\gamma = 0\}},$$

i.e.,  $a, b \in \text{ERV}_{\gamma,1}$ . Furthermore, if functions  $\tilde{a} > 0$  and  $\tilde{b} \in \mathbb{R}$  on  $(0, \infty)$  are *asymptotically equivalent* to  $a, b$ , i.e., they satisfy

$$\frac{\tilde{a}(t)}{a(t)} \longrightarrow 1 \quad \text{and} \quad \frac{\tilde{b}(t) - b(t)}{a(t)} \longrightarrow 0 \quad \text{as } t \rightarrow \infty,$$

then (2.5) and (2.7) hold with  $a, b$  replaced by  $\tilde{a}, \tilde{b}$ . It follows that (2.5) is equivalent to  $t \mathbb{P}[b^{\leftarrow}(Y) > ty] \rightarrow y^{-1}$  for  $y > 0$ , i.e.,  $1 - F_{b^{\leftarrow}(Y)} \in \text{RV}_{-1}$ . This is known as standardization (see [16, Chapter 5]). We say that  $Y^*$  is in the *standardized domain of attraction*, and write  $F_{Y^*} \in D(G_1^*)$ , if

$$(2.8) \quad t \mathbb{P}[Y > ty] \longrightarrow y^{-1}, \quad y > 0,$$

a variant of (2.5) for  $\gamma = 1$ .

**2.3. The Conditional Extreme Value (CEV) Model.** Denote by  $\overline{\mathbb{E}}_\gamma$  the closure on the right of the interval  $\mathbb{E}_\gamma$ . A bivariate random vector  $(X, Y)$  on  $\mathbb{R}^2$  follows a *conditional extreme value model* (CEVM) if there exists a measure  $\mu \in \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma)$ , and functions  $a(t), \alpha(t) > 0$ ,  $b(t), \beta(t) \in \mathbb{R}$ , such that, as  $t \rightarrow \infty$ ,

$$(2.9) \quad t \mathbb{P} \left[ \left( \frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)} \right) \in \cdot \right] \xrightarrow{v} \mu(\cdot) \quad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma),$$

and where  $\mu$  satisfies the *conditional non-degeneracy* conditions: for each  $y \in \mathbb{E}_\gamma$ ,

$$(2.10) \quad \begin{aligned} \mu([-\infty, x] \times (y, \infty]) &\text{ is not a degenerate distribution in } x; \\ \mu(\{\infty\} \times (y, \infty]) &= 0. \end{aligned}$$

It is convenient to choose the normalization such that

$$(2.11) \quad H(x) := \mu([-\infty, x] \times (0, \infty]) \text{ is a probability distribution on } [-\infty, \infty].$$

See [5, 10] for details and [11] for background.

Some remarks: By applying the joint convergence (2.9) to rectangles  $[-\infty, \infty] \times (y, \infty]$ , we see that the distribution of  $Y$  is necessarily attracted to  $G_\gamma$  for some  $\gamma$ . Also, an important property is that the functions  $\alpha, \beta$  are ERV for some  $\rho, k \in \mathbb{R}$  [10, Proposition 1]. The limit measure  $\mu$  in (2.9) is a product measure if and only if  $(\rho, k) = (0, 0)$  [10, Proposition 2].

Condition (2.10) is somewhat different from what is given in [5, 10] which failed to preclude mass on the line  $\{\infty\} \times (-\infty, \infty]$  through infinity. Mass on this line invalidates the convergence to types theorem [8, 14] and since the theory in [5, 10] employs convergence of types arguments, we require the second condition in (2.10). Condition (2.9) entails  $Y \in D(G_\gamma)$  and  $\mu([-\infty, x] \times \{\infty\}) = 0$ . Example 3.6 is a case where (2.9) holds for two distinct normalizations, which are not asymptotically equivalent, yielding two distinct limit measures. One limit measure has  $\mu(\{\infty\} \times (y, \infty]) > 0$  and the other has  $\mu(\{\infty\} \times (y, \infty]) = 0$ .

### 3. STANDARD CASE

Let  $(X, Y)$  be a random vector on  $\mathbb{R}^2$ , with dependence specified by a transition kernel  $K$ :

$$\mathbb{P}[X \in A \mid Y = y] = K(y, \cdot) \quad y \in \mathbb{R}.$$

$K(y, A)$  is a measure in the second variable  $A$  and measurable in  $y$  for each fixed  $A$ . We show if the distribution of  $Y$  is in an extremal domain of attraction, and  $K$  belongs to the domain of attraction of a probability distribution  $G$  (a notion to be defined precisely), then  $(X, Y)$  follows a CEVM.

We begin with the *standard case* which means that  $(X, Y) \in [0, \infty)^2$ , and  $F_Y \in D(G_1^*)$ ,

$$(3.1) \quad tF_Y(t \cdot) \xrightarrow{v} \nu_1(\cdot) \quad \text{in } \mathbb{M}_+(0, \infty] \quad \text{as } t \rightarrow \infty,$$

where  $\nu_1(x, \infty] = x^{-1}$ ,  $x > 0$  (a formulation equivalent to (2.8)) and  $K \in D(G)$  meaning

$$(3.2) \quad K(t, t \cdot) \Rightarrow G(\cdot) \quad \text{on } [0, \infty].$$

In what follows,  $\xi$  will always be a random variable with distribution  $G$ .

**3.1. Standard CEVM Properties.** Conditions (3.1) and (3.2) imply  $(X, Y)$  follows a CEVM, provided  $G \neq \epsilon_0$ , i.e., unit mass at  $\{0\}$ .

**Theorem 3.1.** *Suppose that the joint distribution of the random vector  $(X, Y)$  on  $[0, \infty)^2$  satisfies (3.1) and (3.2), where  $G$  is a probability distribution on  $[0, \infty)$ . Then*

$$(3.3) \quad t\mathbb{P}[(X, Y) \in t \cdot] \xrightarrow{v} \mu(\cdot) \quad \text{in } \mathbb{M}_+([0, \infty] \times (0, \infty]),$$

with limit measure  $\mu$  given for  $x, y > 0$ ,  $\xi \sim G$  by

$$(3.4) \quad \mu([0, x] \times (y, \infty]) = \int_y^\infty G(x/u) \nu_1(du) = \frac{1}{x} \int_0^{x/y} G(u) du = y^{-1} \mathbb{P}[\xi \leq \frac{x}{y}] - x^{-1} \mathbb{E} \xi \mathbf{1}_{\{\xi \leq x/y\}}.$$

Furthermore,  $\mu$  satisfies the conditional non-degeneracy conditions (2.10) provided  $G \neq \epsilon_0$ .

**Proof.** The convergence (3.3) is special case of Proposition 5.1 of [18]; it is an elaboration of the continuous mapping theorem. From (3.4),  $\mu([0, x] \times (y, \infty])$  is continuous in  $x$ , and not constant provided  $G \neq \epsilon_0$ . Also, since  $\mu((x, \infty] \times (y, \infty]) = \int_{(y, \infty]} \nu_1(du) \mathbb{P}[\xi > xu^{-1}]$ , that  $\mu(\{\infty\} \times (y, \infty]) = 0$  follows from the fact that  $G(\{\infty\}) = 0$ . Therefore,  $\mu$  satisfies (2.10).  $\square$

3.1.1. *Properties of the limit measure  $\mu$ .* From (3.4) we see  $\mu$  is continuous in  $x$  and  $y$  and if  $G$  has a density, then so does  $\mu$ . Continuity in (3.4) holds even if  $G$  is degenerate, i.e.,  $G = \epsilon_c$  for some  $c > 0$ ; see Example 3.4 (p. 6). Non-degeneracy of  $G$  only becomes relevant in the non-standard case. Moreover,  $\mu$  cannot be a product measure [5, Lemma 3.1].

From (3.4) we also observe that the  $y$ -axis through the origin is assigned mass proportional to  $G(\{0\})$  since  $\mu(\{0\} \times (y, \infty]) = y^{-1}G(\{0\})$ . Mass on vertical slices of space depends on  $E\xi$ , since  $\mu((x, \infty] \times (0, \infty]) = x^{-1}E\xi \leq \infty$ . In terms of conditional distributions, (3.3) implies

$$(3.5) \quad \mathbb{P}[X \leq tx \mid Y > t] \Rightarrow H(x) := \mu([0, x] \times (1, \infty]) = \frac{1}{x} \int_0^x G(u) du.$$

3.1.2. *Extending to a larger cone.* Convergence (3.3) extends to standard regular variation on the larger cone  $[0, \infty]^2 \setminus \{\mathbf{0}\}$ , so that the distribution of  $(X, Y)$  is in a bivariate domain of attraction, if and only if  $F_X \in D(G_1^*)$  as well [5, Proposition 4.1]. In this case,

$$(3.6) \quad t \mathbb{P}[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^c] \longrightarrow \frac{1}{x} \left( 1 + \int_0^{x/y} G(u) du \right),$$

implying that  $E\xi \leq 1$ , and the  $x$ -axis receives mass according to  $\mu((x, \infty] \times \{0\}) = x^{-1}(1 - E\xi)$ .

3.1.3. *Degenerate  $G$ ; asymptotic independence.* If  $G = \epsilon_0$ , then the convergence (3.3) holds with limit measure  $\mu([0, x] \times (y, \infty]) = y^{-1}$  but conditional non-degeneracy (2.10) fails, since all the mass lies on the  $y$ -axis, so  $(X, Y)$  does not follow a standard CEVM. This is in fact a manifestation of asymptotic independence. Indeed,

$$\mathbb{P}[X > tx \mid Y > t] \rightarrow 0$$

for any  $x$ , so, given that  $Y$  is extreme (exceeding the threshold  $u(t) = t$ ), it is very unlikely to observe  $X$  to be similarly extreme. If the joint distribution of  $(X, Y)$  is regularly varying on the larger cone  $[0, \infty]^2 \setminus \{\mathbf{0}\}$ , then

$$t \mathbb{P}[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^c] \longrightarrow x^{-1} + y^{-1},$$

which means that  $X$  and  $Y$  are asymptotically independent in the usual sense [10, Section 5]. In this case,  $(X, Y)$  does not follow a standard CEVM because of degeneracy, although a CEVM may hold if  $X$  is normalized differently; see Section 4.

This suggests viewing the parameter  $G(\{0\})$  as a measure of asymptotic dependence from  $Y$  to  $X$ . For example, given  $Y$ , we could write  $X$  as a mixture

$$(3.7) \quad X = WX_0 + (1 - W)X_1,$$

where  $X_0$  and  $Y$  are asymptotically independent,  $X_1$  and  $Y$  are asymptotically dependent, and  $W \sim \text{Bernoulli}(G(\{0\}))$ . This is suggested by the canonical form of the update function representation of  $K \in D(G)$  [18, Section 2.3]. Asymptotic dependence in the reverse direction, given large  $X$ , would then be quantified by  $1 - E\xi$  if appropriate. The latter phenomenon is hinted at by Segers [20] in his definition of the “back-and-forth tail chain” to approximate stationary Markov chains.

**3.2. Examples.** Examples illuminate properties of the CEVM based on Markov kernels as in (3.2). First, as in [5, Example 8], given any distribution  $G$  on  $[0, \infty)$ , we construct a CEVM whose limit measure  $\mu$  is built on  $G$  as in (3.4).

**Example 3.1.** Take  $G$  to be any probability distribution on  $[0, \infty)$  not concentrating at 0. Let  $Y \sim \text{Pareto}(1)$  on  $[1, \infty)$ ,  $\xi \sim G$ , independent of  $Y$ , and put  $X = \xi Y$ . A version of the conditional distribution is

$$K(y, \cdot) = \mathbb{P}[X \in \cdot \mid Y = y] = \mathbb{P}[\xi Y \in \cdot \mid Y = y] = G(y^{-1} \cdot),$$

and  $K$  satisfies (3.2) and in fact  $K(t, t \cdot) = G(\cdot)$ . Consequently,  $(X, Y)$  follows a standard CEVM with limit measure as in (3.4). In fact, for  $x, y > 0$ , we have

$$\begin{aligned} \mathbb{P}[X \leq x, Y > y] &= \int_{(y, \infty]} K(u, [0, x]) P[Y \in du] \\ &= \int_{y \vee 1}^{\infty} \mathbb{P}[\xi \leq xu^{-1}] u^{-2} du = \frac{1}{x} \int_0^{x \wedge \frac{x}{y}} G(u) du. \end{aligned}$$

Furthermore,  $(X, Y)$  belong to a standard bivariate domain of attraction (3.6) iff  $F_X \in D(G_1^*)$  as well. The marginal distribution of  $X = \xi Y$  is

$$F_X(x) = \frac{1}{x} \int_0^x G(u) du = H(x),$$

(from (3.5)) which has density  $f_X(x) = x^{-1}(G(x) - H(x))$  for  $x \geq 0$ . Since

$$\lim_{t \rightarrow \infty} t \mathbb{P}[X > tx] = \lim_{t \rightarrow \infty} \frac{1}{x} \int_0^{tx} \mathbb{P}[\xi > u] du = x^{-1} \mathbb{E} \xi \quad (\leq \infty),$$

$(X, Y)$  belongs to the standard domain of attraction iff  $\mathbb{E} \xi = 1$ .  $\square$

Using the Example 3.1 recipe, we explore the CEVM in a variety of special cases.

**Example 3.2.** Choose  $\xi \sim \text{Exp}(\lambda)$  and we have  $X = \lambda^{-1} Y E$ , where  $E \sim \text{Exp}(1)$ . The limit measure is

$$\mu([0, x] \times (y, \infty]) = \frac{1}{x} \int_0^{x/y} (1 - e^{-\lambda u}) du = \frac{1}{y} - \frac{1}{\lambda x} + \frac{e^{-\lambda x/y}}{\lambda x},$$

and the marginal distribution of  $X$  is  $F_X(x) = 1 - (\lambda x)^{-1}(1 - e^{-\lambda x})$  with density  $f(x) = \lambda^{-1} x^{-2}(1 - e^{-\lambda x}) - x^{-1} e^{-\lambda x}$ , and  $F_X$  satisfies (3.1) iff  $\lambda = 1$ .  $\square$

Next, we suppose  $\xi$  is heavy-tailed.

**Example 3.3.** For  $\alpha > 0$  let  $\xi \sim \text{Pareto}(\alpha)$  so  $1 - G(x) =: \bar{G}(x) = x^{-\alpha}$ ,  $x \geq 1$ . The limit measure assigns no mass to  $\{(x, y) : 0 \leq x \leq y\}$ , and for  $x > y > 0$ ,

$$\mu([0, x] \times (y, \infty]) = \begin{cases} \frac{1}{y} - \left(\frac{\alpha}{\alpha-1}\right) \frac{1}{x} + \frac{y^{\alpha-1}}{x^{\alpha}(\alpha-1)} & \alpha > 1 \\ \frac{1}{y} - \frac{1}{x} - \frac{\log x}{x} + \frac{\log y}{y} & \alpha = 1 \\ \frac{1}{y} + \left(\frac{2-\alpha}{1-\alpha}\right) \frac{1}{x} + \frac{1}{x^{\alpha} y^{1-\alpha} (1-\alpha)} & \alpha < 1. \end{cases}$$

When  $\alpha \leq 1$ ,  $\mathbb{E} \xi = \infty$  and  $\mu((x, \infty] \times (y, \infty]) = y^{-1} - \mu([0, x] \times (y, \infty]) \rightarrow \infty$  as  $y \downarrow 0$ .  $\square$

It is also possible that  $G$  is discrete, although the CEVM limit measure  $\mu$  remains continuous.

**Example 3.4.** Suppose  $\xi$  has discrete distribution  $\mathbb{P}[\xi = k] = a_k$ ,  $k = 0, 1, \dots$ . In this case, the limit measure is given by

$$\mu([0, x] \times (y, \infty]) = \frac{1}{x} \int_0^{x/y} \left( \sum_{k=0}^{\lfloor u \rfloor} a_k \right) du = \sum_{k=0}^{\lfloor x/y \rfloor} a_k (y^{-1} - kx^{-1}),$$

which is continuous in  $x$  and  $y$ , and  $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} a_k (1 - kx^{-1})$ . In particular, if  $\mathbb{P}[\xi = c] = 1$  for some  $c > 0$ , we obtain

$$\mu([0, x] \times (y, \infty]) = (y^{-1} - cx^{-1}) \mathbf{1}_{\{x > cy > 0\}}.$$

The conditional non-degeneracy conditions (2.10) are satisfied even though  $G$  is degenerate.  $\square$

The final example shows how  $G$  reflects asymptotic independence between  $X$  and  $Y$ .

**Example 3.5.** Consider  $Y \sim \text{Pareto}(1)$ , and  $Z$  independent of  $Y$  such that  $\mathbb{P}[Z < \infty] = 1$ . Take  $X = Y \vee Z$ . Given  $Y$  is extreme, it is unlikely that  $Z$  is as extreme as  $Y$  since they are independent. We have

$$K(y, [0, x]) = \mathbb{P}[Y \leq x, Z \leq x | Y = y] = \mathbb{P}[Z \leq x] \mathbf{1}_{\{x \geq y\}},$$

and so

$$K(t, t[0, x]) = \mathbb{P}[Z \leq tx] \mathbf{1}_{\{x \geq 1\}} \longrightarrow \mathbf{1}_{\{x \geq 1\}} = \epsilon_1([0, x]) = G([0, x]).$$

As in the previous example, the limit measure is

$$\mu([0, x] \times (y, \infty)) = (y^{-1} - x^{-1}) \mathbf{1}_{\{x > y > 0\}}.$$

On the other hand, if  $X' = Y \wedge Z$  then when  $Y$  is large, it is likely  $X' = Z$ , so  $X'$  should be asymptotically independent of  $Y$ . More precisely,

$$K(y, (x, \infty]) = \mathbb{P}[Y > x, Z > x | Y = y] = \mathbb{P}[Z > x] \mathbf{1}_{\{y > x\}},$$

from which

$$K(t, t(x, \infty]) = \mathbb{P}[Z > tx] \mathbf{1}_{\{x < 1\}} \longrightarrow 0$$

for  $x > 0$ . Therefore,  $G = \epsilon_0$ , and the conditional non-degeneracy conditions fail.  $\square$

**3.3. Counter-examples.** As expected, the converse to Theorem 3.1 can fail. If  $(X, Y)$  follows a non-degenerate CEVM as in (3.3), and  $K$  is a specific version of the conditional distribution  $\mathbb{P}[X \in \cdot | Y = y]$ , it does not necessarily follow that there exists a distribution  $G$  such that (3.2) holds. The failure of (3.2) can happen in two ways. There may exist a probability distribution  $G$  on  $[0, \infty]$  satisfying (3.2) with  $G(\{\infty\}) > 0$  or it may be possible to obtain two distinct limit distributions down different subsequences  $\{t_n\}$  and  $\{t'_n\}$ .

Example 3.6 where  $G(\{\infty\}) > 0$  emphasizes the importance of assuming  $\mu(\{\infty\} \times (y, \infty]) = 0$ .

**Example 3.6.** As usual, take  $Y \sim \text{Pareto}(1)$  and suppose that

$$X = WY + (1 - W)Y^2,$$

where  $W \sim \text{Bernoulli}(p)$  independent of  $Y$ . Then

$$K(y, \cdot) = \mathbb{P}[X \in \cdot | Y = y] = p\epsilon_y + (1 - p)\epsilon_{y^2},$$

so

$$K(t, t \cdot) = p\epsilon_1 + (1 - p)\epsilon_t \Rightarrow p\epsilon_1 + (1 - p)\epsilon_\infty = G \quad \text{on } [0, \infty].$$

Indeed, for  $0 \leq x < \infty$ ,

$$K(t, t[0, x]) = p\epsilon_1([0, x]) + (1 - p)\epsilon_t([0, x]) \longrightarrow p\epsilon_1([0, x])$$

showing that  $G(\{\infty\}) = 1 - p$ .

On the other hand, for  $x, y > 0$ ,

$$\begin{aligned} \mathbb{P}[X \leq x, Y > y] &= p\mathbb{P}[Y \leq x, Y > y] + (1 - p)\mathbb{P}[Y^2 \leq x, Y > y] \\ &= p\left[\frac{1}{(y \vee 1)} - \frac{1}{x}\right] \mathbf{1}_{\{x \geq (y \vee 1)\}} + (1 - p)\left[\frac{1}{(y \vee 1)} - \frac{1}{\sqrt{x}}\right] \mathbf{1}_{\{x \geq (y \vee 1)^2\}}, \end{aligned}$$

so for  $t$  sufficiently large,

$$\begin{aligned} t\mathbb{P}[X \leq tx, Y > ty] &= p\left[\frac{1}{y} - \frac{1}{x}\right] \mathbf{1}_{\{x \geq y\}} + (1 - p)\left[\frac{1}{y} - \frac{\sqrt{t}}{\sqrt{x}}\right] \mathbf{1}_{\{\sqrt{x}/\sqrt{t} \geq y\}} \\ (3.8) \quad &\longrightarrow p\left[\frac{1}{y} - \frac{1}{x}\right] \mathbf{1}_{\{x \geq y\}} = \mu([0, x] \times (y, \infty)). \end{aligned}$$



The measure  $\mu$  assigns positive mass to  $\{\infty\} \times (y, \infty]$  since

$$\mu((x, \infty] \times (y, \infty]) = y^{-1} - \mu([0, x] \times (y, \infty]) = \frac{1}{y} \mathbf{1}_{\{x < y\}} + \left[ \frac{1-p}{y} + \frac{p}{x} \right] \mathbf{1}_{\{x \geq y\}},$$

and thus  $\mu(\{\infty\} \times (y, \infty]) = (1-p)y^{-1}$ . Therefore,  $\mu$  does not satisfy (2.10).

Under a different normalization, we obtain a proper limit  $G$ . Indeed, note that

$$K(t, t^2 \cdot) = p\epsilon_{t^{-1}} + (1-p)\epsilon_1 \Rightarrow p\epsilon_0 + (1-p)\epsilon_1 \sim \text{Bernoulli}(1-p),$$

and hence,

$$\begin{aligned} t \mathbb{P}[X \leq t^2 x, Y > t \cdot y] &= p(y^{-1} - (tx)^{-1}) \mathbf{1}_{\{x \geq y/t\}} + (1-p)(y^{-1} - x^{-1/2}) \mathbf{1}_{\{x \geq y\}} \\ &\longrightarrow py^{-1} + (1-p)(y^{-1} - x^{-1/2}) \mathbf{1}_{\{x \geq y\}}. \end{aligned}$$

This limit does satisfy (2.10).  $\square$

Without condition (2.10), the convergence of types theorem fails and it is possible to obtain different CEV limits under different normalizations. From (3.4),  $\mu(\{\infty\} \times (y, \infty]) = G(\{\infty\})y^{-1}$  and excluding defective distributions in Theorem 3.1 avoids cases like the previous one.

Here is an example of a CEVM where the normalized kernel  $K$  does not have a unique limit.

**Example 3.7.** Suppose  $Y \sim \text{Pareto}(1)$ , and define  $X$  by

$$X = WY + (1-W)2Y \mathbf{1}_{\{Y \in [0, \infty) \setminus \mathbb{N}\}}$$

where  $W \sim \text{Bernoulli}(p)$  independent of  $Y$ . In other words, given  $Y = y$ ,  $X$  takes the value  $y$  or  $2y$  according to a coin flip, unless  $y$  is an integer, in which case  $X$  will be either  $y$  or 0. The CEVM holds for  $(X, Y)$ . Since  $\mathbb{P}[Y \in \mathbb{N}] = 0$ , we have

$$\begin{aligned} \mathbb{P}[X \leq x, Y > y] &= \mathbb{P}[X \leq x, Y > y, Y \in [0, \infty) \setminus \mathbb{N}] \\ &= p \mathbb{P}(Y \leq x, Y > y) + (1-p) \mathbb{P}(2Y \leq x, Y > y) \\ &= p(y^{-1} - x^{-1}) \mathbf{1}_{\{x \geq y\}} + (1-p)(y^{-1} - 2x^{-1}) \mathbf{1}_{\{x \geq 2y\}}, \end{aligned}$$

and  $t \mathbb{P}[X \leq tx, Y > ty] = \mathbb{P}[X \leq x, Y > y]$ , which satisfies (2.11) and the requirement that  $\mu((\cdot) \times (y, \infty])$  not be degenerate for any  $y$ . However, the conditional distribution of  $X$  given  $Y$  is

$$K(y, \cdot) = \begin{cases} p\epsilon_y + (1-p)\epsilon_0 & y \in \mathbb{N} \\ p\epsilon_y + (1-p)\epsilon_{2y} & y \in [0, \infty) \setminus \mathbb{N} \end{cases},$$

so

$$K(t, t \cdot) = \begin{cases} p\epsilon_1 + (1-p)\epsilon_0 & t \in \mathbb{N} \\ p\epsilon_1 + (1-p)\epsilon_2 & t \in [0, \infty) \setminus \mathbb{N} \end{cases}.$$

We obtain different limits along the sequences  $t_n = n$  and  $t'_n = n/2$  and  $K(t, t \cdot)$  does not converge.  $\square$

The technical difficulty highlighted in Example 3.7 is that conditional distributions of the form  $\mathbb{P}[X \in \cdot \mid Y = y]$  are only specified up to sets of  $\mathbb{P}[Y \in \cdot]$ -measure zero. If  $Y$  is absolutely continuous, we can alter the conditional probability for a countable number of  $y$  without affecting the joint distribution. Consequently, constructing a convergence theory based on conditional distributions requires care. The best one can do is fix a version of the kernel or, if circumstances allow, choose a version of the kernel with some claim to naturalness based on smoothness. This is the reason the approach in [5, 10] is based on vague convergence of measures rather than convergence of conditional distributions as in [11].



4. GENERAL NORMALIZATION FOR  $X$ 

The CEVM allows different normalizations for  $X$  and  $Y$ , as in (2.9), but the formulation  $K \in D(G)$  in (3.2), imposes the same normalization for both. We now allow general linear normalizations of  $X$  in the kernel condition, continuing to assume condition (3.1) that  $F_Y \in D(G_1^*)$ .

We will assume the following generalization of (3.2): there exist scaling and centering functions  $\alpha(t) > 0$ ,  $\beta(t) \in \mathbb{R}$  and a non-degenerate probability distribution  $G$  on  $[-\infty, \infty)$ , such that

$$(4.1) \quad K(t, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G([-\infty, x]) \quad \text{on } [-\infty, \infty].$$

**4.1. CEVM Properties.** Consider the decomposition

$$(4.2) \quad t \mathbb{P} \left[ \frac{X - \beta(t)}{\alpha(t)} \leq x, Y > ty \right] = \int_{(y, \infty]} t \mathbb{P}[Y \in tdu] K(tu, [-\infty, \alpha(t)x + \beta(t)]).$$

By a variant of the continuous mapping theorem ([18, Lemma 8.2], the integrals converge provided  $K(tu(t), [-\infty, \alpha(t)x + \beta(t)]) \rightarrow \varphi_x(u)$  whenever  $u(t) \rightarrow u > 0$ . Proposition 4.1 discusses when this happens.

Given  $\rho, k \in \mathbb{R}$ , define the *generalized tail kernel* associated with a distribution  $G$  on  $[-\infty, \infty]$  as the transition function  $\kappa_G : (0, \infty) \times \mathcal{B}[-\infty, \infty] \rightarrow [0, 1]$  given by

$$(4.3) \quad \kappa_G(y, A) = G(y^{-\rho}[A - \psi(y)]),$$

where  $\psi$  is specified in (2.2) (p. 2). Note that  $\kappa_G$  describes transitions between two different spaces. Since  $\psi$  satisfies  $\psi(uy) = u^\rho \psi(y) + \psi(u)$ , a kernel  $\kappa$  has the form (4.3) iff

$$(4.4) \quad \kappa(uy, A) = \kappa(y, u^{-\rho}[A - \psi(u)]).$$

**Proposition 4.1.** *Let  $K : (0, \infty) \times \mathcal{B}[-\infty, \infty] \rightarrow [0, 1]$  be a transition function satisfying (4.1) with  $G$  is non-degenerate. There exists a family of non-degenerate probability distributions  $\{G_u : 0 < u < \infty\}$  on  $[-\infty, \infty)$  such that for  $0 < u < \infty$ ,*

$$(4.5) \quad K(tu, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G_u([-\infty, x]) \quad \text{on } [-\infty, \infty],$$

as  $t \rightarrow \infty$  if and only if  $\alpha, \beta \in \text{ERV}_{\rho, k}$  as in (2.1) (p. 2). In this case,  $G_1 = G$ , and

$$(4.6) \quad K(tu_t, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow \kappa_G(u, [-\infty, x]) \quad \text{on } [-\infty, \infty]$$

whenever  $u_t = u(t) \rightarrow u \in (0, \infty)$ ; i.e., the limit is a transition function of the form (4.3), where  $\rho, k$  are the ERV parameters of  $\alpha, \beta$ .

**Proof.** Assume first that  $\alpha, \beta \in \text{ERV}_{\rho, k}$  and define

$$h_t(y; u) = \frac{\alpha(tu)}{\alpha(t)}y + \frac{\beta(tu) - \beta(t)}{\alpha(t)},$$

so that by (2.1),  $h_t(y_t; u) \rightarrow h(y; u) = u^\rho y + \psi(u)$  whenever  $y_t \rightarrow y \in \mathbb{R}$ . For  $u > 0$ ,

$$K(tu, [-\infty, \alpha(t)x + \beta(t)]) = K(tu, \alpha(tu)\{h_t^{-1}(\cdot; u)[-\infty, x]\} + \beta(tu)).$$

Applying the second continuous mapping theorem ([1], [18, Lemma 8.1]) to (4.1), we have

$$K(tu, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow (G \circ h^{-1}(\cdot; u))([-\infty, x]) = G([-\infty, (x - \psi(u))/u^\rho]).$$

Hence, (4.5) holds with  $G_u(\cdot) = \kappa_G(u, \cdot)$  and  $G_1 = G$ . Furthermore, we have  $h_t(x_t; u_t) \rightarrow h(x; u)$  whenever  $u_t \rightarrow u > 0$ , establishing (4.6).

For the converse, we employ convergence of types. Denote by  $H_t(\cdot)$  the distribution  $K(t, \cdot)$ . Then, on the one hand, we have  $H_t([-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G_1([-\infty, x])$ . On the other hand, fixing  $c > 0$ , we have

$$H_t(\alpha(tc)x + \beta(tc)) = K((tc)c^{-1}, [-\infty, \alpha(tc)x + \beta(tc)]) \Rightarrow G_{c^{-1}}([-\infty, x]).$$

Convergence of types yields that  $\alpha, \beta \in \text{ERV}_{\rho, k}$ , and

$$G_{c^{-1}}([-\infty, x]) = G_1([-\infty, c^\rho x + \psi(c)]),$$

with  $\psi$  as in (2.2). Using the identity (2.3), we find that  $G_u$  has the form (4.3), with  $G = G_1$ .  $\square$

Starting from kernel convergence (4.1), Proposition 4.1 implies that  $\alpha, \beta$  being ERV is necessary and sufficient for obtaining a CEVM. Unlike Section 3, here we need  $G$  to be non-degenerate in order to apply the convergence of types theorem.

**Theorem 4.1.** *Suppose  $(X, Y)$  is a random vector on  $\mathbb{R} \times [0, \infty)$  and (3.1) holds. Assume (4.1) holds for non-degenerate limit distribution  $G$  on  $[-\infty, \infty)$  and scaling and centering functions  $\alpha(t) > 0$  and  $\beta(t) \in \mathbb{R}$ . As  $t \rightarrow \infty$ ,*

$$(4.7) \quad t \mathbb{P} \left[ \left( \frac{X - \beta(t)}{\alpha(t)}, \frac{Y}{t} \right) \in \cdot \right] \xrightarrow{v} \mu(\cdot) \neq 0 \quad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty])$$

where  $\mu$  satisfies (2.10), if and only if  $\alpha, \beta \in \text{ERV}_{\rho, k}$ . In this case,  $\mu$  is specified by

$$(4.8) \quad \mu([-\infty, x] \times (y, \infty]) = \int_{(y, \infty]} \nu_1(du) G(u^{-\rho}(x - \psi(u))), \quad x \in \mathbb{R}, \quad y > 0,$$

with  $\psi$  as in (2.2) and  $\nu_1(du) = u^{-2}du$ ,  $u > 0$ . Expression (4.8) is continuous in  $x$  and  $y$  if  $(\rho, k) \neq (0, 0)$ , or if  $G$  is continuous.

**Proof.** The convergence (4.7) to a limit  $\mu$  satisfying (2.10) implies  $\alpha, \beta \in \text{ERV}$  [10, Proposition 1]. Conversely, if  $\alpha, \beta \in \text{ERV}_{\rho, k}$ , then the convergence (4.7) follows from a variant of the continuous mapping theorem ([18, Lemma 8.4]) in light of (4.6), yielding the limit in (4.8). We check that  $\mu([-\infty, x] \times (y, \infty])$  is continuous when  $(\rho, k) \neq (0, 0)$  by applying dominated convergence: if  $x_n \rightarrow x$ , then

$$G(u^{-\rho}(x_n - \psi(u))) \rightarrow G(u^{-\rho}(x - \psi(u)))$$

for all except a countable number of  $u$  corresponding to discontinuities of the distribution function. Continuity in  $y$  is clear. Also, if  $(\rho, k) = (0, 0)$ , then  $\mu([-\infty, x] \times (y, \infty]) = y^{-1}G(x)$ , which is continuous if  $G$  is. In either case,  $\mu([-\infty, x] \times (y, \infty])$  is non-degenerate in  $x$  because  $G$  is non-degenerate. Finally,  $\mu(\{\infty\} \times (y, \infty]) = y^{-1}G(\{\infty\}) = 0$ . Therefore,  $\mu$  satisfies (2.10).  $\square$

Changing variables  $u \mapsto 1/u$  in (4.8), the limit measure is

$$(4.9) \quad \mu([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} G(u^\rho x + \psi(u)) du,$$

where

$$u^\rho x + \psi(u) = \begin{cases} u^\rho(x + k\rho^{-1}) - k\rho^{-1} & \rho \neq 0 \\ x + k \log u & \rho = 0 \end{cases}.$$

Changing variables, we obtain the following expressions for  $\mu$  according to  $(\rho, k)$ :

$$(4.10) \quad \mu([-\infty, x] \times (y, \infty]) = \begin{cases} \frac{1}{\rho|x + k\rho^{-1}|^{1/\rho}} \int_0^{|x + k\rho^{-1}|y^{-\rho}} u^{(1-\rho)/\rho} G(u \operatorname{sgn}(x + k\rho^{-1}) - k\rho^{-1}) du & \rho \neq 0 \\ \frac{1}{|k|e^{x/k}} \int_{-\infty}^{x \operatorname{sgn}(k) - |k| \log y} e^{u/|k|} G(u \operatorname{sgn}(k)) du & \rho = 0, \quad k \neq 0 \\ y^{-1}G(x) & \rho = 0, \quad k = 0 \end{cases}.$$

Here  $\text{sgn}(v) = v/|v| \mathbf{1}_{\{v \neq 0\}}$ , and we read the measure as  $y^{-1}G(-k\rho^{-1})$  when  $x = -k\rho^{-1}$  for the case  $\rho \neq 0$ . Continuity in  $x$  and  $y$  when  $(\rho, k) \neq (0, 0)$  is apparent from the above expressions.

We give an example where  $K$  satisfies (4.1), but (4.7) fails because  $\alpha, \beta$  are not ERV.

**Example 4.1.** Consider  $Y \sim \text{Pareto}(1)$  and  $U \sim \text{Uniform}(0, 1)$ , independent of  $Y$ . Put  $X = Ue^Y$ . Then

$$K(y, [0, x]) = \mathbb{P}[X \leq x | Y = y] = \mathbb{P}[U \leq xe^{-y}] = xe^{-y} \wedge 1.$$

Polynomial scaling is not strong enough to give an informative limit, since for any  $\rho > 0$ ,

$$K(t, t^\rho[0, x]) = x^\rho t^\rho e^{-t} \wedge 1 \rightarrow 0.$$

The appropriate normalization is exponential  $(\alpha(t), \beta(t)) = (e^t, 0)$ , which is not ERV:

$$K(t, \alpha(t)[0, x]) = xe^t e^{-t} \wedge 1 \rightarrow x \wedge 1 = G(x),$$

and up to asymptotic equivalence, by the convergence to types theorem, this the only normalization yielding a non-degenerate limit. Since  $(\alpha(t), \beta(t))$  is not ERV, Theorem 4.1 claims that  $(X, Y)$  cannot follow a CEVM. To verify this *ab initio*, consider  $y > 0$  and  $t$  so large that  $ty > 1$ :

$$\begin{aligned} t \mathbb{P}[X \leq \alpha(t)x, Y > ty] &= t \mathbb{P}[Ue^Y \leq e^t x, Y > ty] = \int_y^\infty u^{-2} (xe^{-t(u-1)} \wedge 1) du \\ &= \int_{((y \vee 1), \infty)} u^{-2} du \{xe^{-t(u-1)} \wedge 1\} + \mathbf{1}_{\{y < 1\}} \int_{(y, 1]} u^{-2} du \{xe^{-t(u-1)} \wedge 1\}. \end{aligned}$$

The first integral in the previous sum is bounded by  $xy^{-1}e^{-t(y-1)} \rightarrow 0$ . If  $y \leq 1$ , the second integral approaches  $\nu_1(y, 1] = y^{-1} - 1$ . Therefore, the limit is degenerate in  $x$ , violating conditional non-degeneracy (2.10). This is not repaired by using ERV normalization since if  $\tilde{\alpha}, \tilde{\beta}$  are ERV, then

$$\begin{aligned} t \mathbb{P}[X \leq \tilde{\alpha}(t)x + \tilde{\beta}(t), Y > ty] &= \int_{(y, \infty)} u^{-2} du K(tu, [0, \tilde{\alpha}(t)x + \tilde{\beta}(t)]) \\ &= \int_{(y, \infty)} u^{-2} du \{e^{-tu}(\tilde{\alpha}(t)x + \tilde{\beta}(t)) \wedge 1\} \rightarrow 0, \end{aligned}$$

which follows from the asymptotic properties of ERV functions (see Section 2.1 (p. 2)).  $\square$

**4.2. Standardization of  $X$ .** In certain cases, it is possible to standardize the  $X$  variable [5, Section 3.2].

**4.2.1. Standardization functions.** Denote by  $x^*$  and  $x_*$  the upper and lower endpoints of the distribution of  $X$  respectively, i.e.,

$$x^* = \sup\{x : F_X(x) < 1\} \quad \text{and} \quad x_* = \inf\{x : F_X(x) > 0\}.$$

Call  $f : (0, \infty) \mapsto (x_*, x^*)$  a *standardization function* if  $f$  is monotone and  $\lim_{x \rightarrow \infty} f(x) = x^*$  if  $f$  is non-decreasing and  $\lim_{x \rightarrow \infty} f(x) = x_*$  if  $f$  is non-increasing. As in [5, Section 3], we standardize with such functions. For the purpose of this section, extend the definition of  $f^{\leftarrow} : (x_*, x^*) \mapsto (0, \infty)$  in order to invert right-continuous monotone functions which are either increasing or decreasing. Define

$$f^{\leftarrow}(x) = \begin{cases} \inf\{y : f(y) \geq x\} & \text{if } f \text{ is non-decreasing} \\ \inf\{y : f(y) \leq x\} & \text{if } f \text{ is non-increasing} \end{cases}.$$

Note that  $f^{\leftarrow}$  is left-continuous for  $f$  non-decreasing and right-continuous for  $f$  non-increasing. The main property we shall be using is that

$$\begin{cases} f^{\leftarrow}(x) \leq y \iff x \leq f(y) & f \text{ non-decreasing} \\ f^{\leftarrow}(x) \leq y \iff x \geq f(y) & f \text{ non-increasing} \end{cases}.$$

The distinction between the two cases is a technicality which should not cause confusion in the following discussion. Also, say that a monotone function  $f$  has two *points of change* if there exist  $x_1 < x_2 < x_3$  such that  $f(x_1) < f(x_2) < f(x_3)$  for  $f$  non-decreasing, and with the opposite inequalities in the non-increasing case.

If the pair  $(X, Y)$  satisfies (4.7) for some  $\alpha > 0$  and  $\beta$ , then we say  $(X, Y)$  can be *standardized* if there exists a standardization function  $f$  and a non-null Radon measure  $\mu^*$  such that

$$(4.11) \quad t \mathbf{P} [t^{-1}(f^{\leftarrow}(X), Y) \in \cdot] \xrightarrow{v} \mu^*(\cdot) \quad \text{in } \mathbb{M}_+([0, \infty] \times (0, \infty)).$$

If the limit  $\mu$  in (4.7) satisfies the conditional non-degeneracy conditions (2.10), then standardization is possible if and only if  $(\rho, k) \neq (0, 0)$ ; i.e.,  $\mu$  is not a product.

**4.2.2. Characterizing standardization functions.** What property does  $f$  need to be a standardization function satisfying (4.11)?

**Proposition 4.2.** *Suppose  $(X, Y)$  follow a CEVM so that (4.7) holds with  $\mu$  satisfying the conditional non-degeneracy conditions (2.10). Assume  $(\rho, k) \neq (0, 0)$ . A function  $f$  standardizes  $(X, Y)$  in the sense of (4.11) where  $\mu^*$  satisfies the conditional non-degeneracy conditions iff*

$$(4.12) \quad \frac{f(tx) - \beta(t)}{\alpha(t)} \rightarrow \varphi(x), \quad x > 0,$$

where  $\varphi$  has at least two points of change. In this case,  $\mu$  and  $\mu^*$  are related by

$$\mu^*([0, x] \times (y, \infty]) = \mu(A_\varphi(x) \times (y, \infty]),$$

where

$$(4.13) \quad A_\varphi(x) = \begin{cases} [-\infty, \varphi(x)] & f \text{ non-decreasing} \\ [\varphi(x), \infty] & f \text{ non-increasing} \end{cases}.$$

It follows that  $\alpha(\cdot), f(\cdot) \in \text{ERV}$ , though not necessarily with the same parameters as  $\alpha, \beta$ . However, depending on the case,  $f$  can be expressed in terms of either  $\beta$  or  $\alpha$  ([4, Proposition 2.3.3]).

**Proof.** Suppose  $f$  is non-decreasing. Then for  $x, y > 0$ , we can write

$$(4.14) \quad t \mathbf{P} \left[ \frac{f^{\leftarrow}(X)}{t} \leq x, \frac{Y}{t} > y \right] = t \mathbf{P} \left[ \frac{X - \beta(t)}{\alpha(t)} \leq \frac{f(tx) - \beta(t)}{\alpha(t)}, \frac{Y}{t} > y \right].$$

If  $f$  satisfies (4.12), then (4.11) holds with

$$\mu^*([0, x] \times (y, \infty]) = \mu([-\infty, \varphi(x)] \times (y, \infty])$$

non-degenerate in  $x$ . On the other hand, if (4.11) holds, then (4.14) implies (4.12), and  $\varphi$  has at least two points of increase because  $\mu^*$  is non-degenerate in  $x$ . The mass at  $\{\infty\}$  condition in (2.10) follows from the fact that  $\lim_{x \rightarrow \infty} \varphi(x) = \infty$  if  $f$  is non-decreasing (see (4.15) below). The case for  $f$  non-increasing is similar, after reversing the inequality for  $X$  on the right-hand side of (4.14).  $\square$

Assuming (4.12) and  $\alpha, \beta \in \text{ERV}_{k, \rho}$ , write

$$\frac{f(tx) - \beta(t)}{\alpha(t)} = \frac{\alpha(tx)}{\alpha(t)} \frac{f(tx) - \beta(tx)}{\alpha(tx)} + \frac{\beta(tx) - \beta(t)}{\alpha(t)}$$

and with  $c = \varphi(1)$ ,  $\varphi$  has the form

$$(4.15) \quad \varphi(x) = \begin{cases} cx^\rho + k\rho^{-1}(x^\rho - 1) & \rho \neq 0 \\ c + k \log x & \rho = 0 \end{cases}.$$

If  $\varphi$  has two points of change, we get the constraint that  $c \neq 0$  if  $\rho \neq 0$ ,  $k = 0$ .

**4.2.3. Kernel convergence and standardization.** Assuming a standardization function exists and that the conditional distribution of  $X$  given  $Y$  satisfies kernel convergence assumption (4.1), we can standardize directly through the Markov kernel. We consider the new kernel  $K_f(y, A) = K(y, f(A)) =: \mathbb{P}[f^\leftarrow(X) \in A | Y = y]$ . The next result may be compared to the formulation in [4, Proposition 2.3.3] for joint distributions rather than Markov kernels.

**Proposition 4.3.** *Suppose the transition function  $K : (0, \infty) \times \mathcal{B}[-\infty, \infty] \rightarrow [0, 1]$  satisfies (4.1) for a probability distribution  $G$  on  $[-\infty, \infty]$ . If  $f$  is a monotone function satisfying (4.12), then the transition function  $K_f : (0, \infty) \times \mathcal{B}[0, \infty] \rightarrow [0, 1]$  defined as*

$$K_f(y, A) = K(y, f(A))$$

*satisfies (3.2),*

$$K_f(t, t[0, x]) \Rightarrow G(A_\varphi(x)) =: G_f([0, x]) \quad \text{on } [0, \infty],$$

*with  $A_\varphi(x)$  as in (4.13). Conversely, if we start with a kernel  $K(y, \cdot)$  satisfying (3.2), for limit probability measure  $G$  on  $[0, \infty)$ , then given ERV functions  $\alpha > 0$ ,  $\beta \in \mathbb{R}$  on  $(0, \infty)$ , if  $f$  is monotone on  $(0, \infty)$  satisfying (4.12), the transition function  $\overline{K}_f : (0, \infty) \times \mathcal{B}[-\infty, \infty] \mapsto [0, 1]$  given by*

$$\overline{K}_f(y, A) = K(y, f^\leftarrow(A))$$

*satisfies (4.1),*

$$\overline{K}_f(t, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G(A_{\varphi^\leftarrow}(x)) =: \overline{G}_f([-\infty, x]) \quad \text{on } f([0, \infty]),$$

*where*

$$A_{\varphi^\leftarrow}(x) = \begin{cases} [0, \varphi^\leftarrow(x)] & f \text{ non-decreasing} \\ [\varphi^\leftarrow(x), \infty] & f \text{ non-increasing} \end{cases}.$$

**Proof.** Assume (4.1) and  $f$  is a non-decreasing function satisfying (4.12). Then,

$$\begin{aligned} K_f(t, t[0, x]) &= K(t, [-\infty, f(tx)]) \\ &= K\left(t, \alpha(t) \left[ -\infty, \frac{f(tx) - \beta(t)}{\alpha(t)} \right] + \beta(t)\right) \Rightarrow G([-\infty, \varphi(x)]). \end{aligned}$$

Conversely, if  $f$  satisfies (4.12) for  $\alpha, \beta \in \text{ERV}$ , then inverting (4.12) yields

$$f^\leftarrow(\alpha(t)x + \beta(t))/t \longrightarrow \varphi^\leftarrow(x), \quad x \in f((0, \infty)).$$

Consequently,

$$\overline{K}_f(t, [-\infty, \alpha(t)x + \beta(t)]) = K(t, t[0, t^{-1}f^\leftarrow(\alpha(t)x + \beta(t))]) \Rightarrow G([0, \varphi^\leftarrow(x)]).$$

The case for non-increasing  $f$  is similar.  $\square$

If  $K$  is a version of the conditional distribution  $\mathbb{P}[X \in \cdot | Y = y]$  satisfying (4.1), where  $\alpha, \beta \in \text{ERV}_{\rho, k}$  with  $(\rho, k) \neq (0, 0)$  and  $F_Y$  is in the standardized domain of attraction, then  $(X, Y)$  follows a CEVM by Theorem 4.1. Furthermore,  $(X, Y)$  can be standardized in the sense of (4.11) [4, Proposition 2.3.3 (1)], and the standardization function  $f$  satisfies (4.12) by Proposition 4.2.

**4.2.4. Moment restrictions.** Section 3.1 considered the standard case and found, in particular, that if  $X$  belongs to the standardized domain of attraction,  $0 \leq \mathbb{E} \xi \leq 1$  where  $\xi \sim G$ . When standardization is possible, a comparable moment restriction occurs provided  $X$  has a distribution in a domain of attraction.

Assume there exist normalizing functions  $c(t) > 0$  and  $d(t) \in \mathbb{R}$  such that

$$(4.16) \quad t \mathbb{P} \left[ \frac{X - d(t)}{c(t)} > x \right] \longrightarrow (1 + \lambda x)^{-1/\lambda} \quad x \in \mathbb{E}_\lambda,$$

implying that  $c, d \in \text{ERV}_{\lambda,1}$  (see Section 2.2, p. 3). If  $(X, Y)$  follow a CEVM and (4.16) holds, then the vector  $(X, Y)$  belongs to a multivariate domain of attraction provided  $\lim_{t \rightarrow \infty} \alpha(t)/c(t) \in [0, \infty)$  [5, Proposition 4.1]. Continuing the theme of assuming kernel convergence, consider the case where (4.1) holds under the same normalization as in (4.16):

$$K(t, [-\infty, c(t)x + d(t)]) \Rightarrow G(x).$$

Then from (2.7), p. 3,  $d$  is a standardization function satisfying (4.12), and

$$\varphi(x) = \begin{cases} \lambda^{-1}(x^\lambda - 1) & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}.$$

Theorem 3.1 gives a standard CEVM for  $(d^\leftarrow(X), Y)$ , and furthermore,  $d^\leftarrow(X)$  belongs to the standardized domain of attraction. Therefore, the distribution  $G$  must satisfy

$$\int_0^\infty \mathbb{P}[\xi > \varphi(x)] dx \leq 1.$$

Depending on  $\lambda$ , this reduces to

$$\begin{cases} \mathbb{E} \xi^{1/\lambda} \mathbf{1}_{\{\xi > 0\}} \leq \lambda^{-1/\lambda} & \lambda > 0 \\ \mathbb{E} (-1/\xi)^{1/|\lambda|} \mathbf{1}_{\{\xi < 0\}} \leq |\lambda|^{1/|\lambda|} & \lambda < 0 \\ \mathbb{E} e^\xi \leq 1 & \lambda = 0 \end{cases}.$$

Thus, we obtain a different condition for each class of extreme value distribution. In the Fréchet case, we have a bound on the  $1/\lambda$ -th moment of the right tail. If the domain of attraction is Weibull, this becomes an integrability condition near 0. Finally, in the Gumbel case, the right tail of  $\xi$  is exponentially bounded, so all right-tail moments exist.

**4.3. Relation to the Heffernan and Tawn Model.** The CEVM of Theorem 4.1 is inspired by the work of Heffernan and Tawn [11]. Where Heffernan and Tawn's model is based on the convergence of conditional distributions as in (4.1), the general CEVM defined in (2.9), (2.10) focuses on limits of joint distributions. Our Theorem 4.1 shows that Heffernan and Tawn's assumption [11, Equation (3.1)] leads to a CEVM provided their normalization functions  $\alpha$  and  $\beta$  are ERV. The fact that they require convergence (4.1) hold at all points  $x$  suggests that they are expecting continuous limits whereas we framed the assumption as weak convergence.

A condition such as (4.1) tacitly assumes a particular version of the conditional distribution. The issue of version cannot be ignored, since Example 3.7 shows that (4.1) holding for one particular version does not imply that it holds for every version. The issue of version is usually handled by smoothness assumptions.

For a non-degenerate CEVM, the functions  $\alpha$  and  $\beta$  are necessarily ERV. Heffernan and Tawn assume a parametric form for these functions. They specify

$$\alpha(y) = b_{|i}(y) := y^{b_{|i}} = y^\rho$$

for some constant  $\rho < 1$  and

$$\beta(y) = a_{|i}(y) := \begin{cases} ay & 0 \leq \rho < 1, \text{ with } a \in [0, 1] \\ c - d \log y & \rho < 0 \text{ with } a = 0, c \in \mathbb{R}, d \in [0, 1] \end{cases}.$$

Although more general models are possible, the form of the ERV limit function  $\psi$  in (2.2) (p. 2) suggests that a parametric approach is indeed reasonable.

5. GENERAL NORMALIZATIONS FOR BOTH  $X$  AND  $Y$ 

So far we assumed that  $Y$  satisfies  $tP[Y > ty] \rightarrow y^{-1}$  for  $y > 0$ . We now extend Theorem 4.1 to the case where  $Y$  belongs to a general domain of attraction:

$$(5.1) \quad tP[Y > a(t)y + b(t)] \rightarrow (1 + \gamma y)^{-1/\gamma} \quad y \in \mathbb{E}_\gamma,$$

and  $\mathbb{E}_\gamma := \{y : 1 + \gamma y > 0\}$ . Assume  $b(t)$  is given by (2.6). Without change, (4.1) may no longer be sufficient to obtain a general CEVM limit (2.9) if  $Y$  requires normalization according to  $a$  and  $b$ .

To relate kernel convergence to the CEVM when (5.1) is the hypothesis, there are two ways to proceed: (i) Assume  $K(a(t)u + b(t), [-\infty, \alpha(t)x + \beta(t)]) \rightarrow \varphi_x(u)$  for  $u > 0$ , and then (2.9) should follow from arguments similar to those in Section 4.1. (ii) Standardize  $Y$  via the transformation  $Y \mapsto b^\leftarrow(Y)$ , use a version of  $P[X \in \cdot | b^\leftarrow(Y) = y] =: K^*(y, \cdot)$ , and rely on (4.1) for  $K^*$ . We show the consistency of these two approaches.

**5.1. Kernel Asymptotics.** The transition function  $K : (-\infty, \infty) \times \mathcal{B}[-\infty, \infty] \mapsto [0, 1]$  is a specific version of the conditional distribution of  $X$  given  $Y$ ,  $K(y, \cdot) = P[X \in \cdot | Y = y]$ . To consider (ii) above, we first express a version the conditional distribution of  $X$  given  $b^\leftarrow(Y)$  in terms of  $K$ .

When the distribution of  $Y$  is not in the Fréchet domain of attraction, the convergence (5.1), where  $b$  is given by (2.6), implies that  $a, b \in \text{ERV}_{\gamma,1}$  for some  $\gamma \in \mathbb{R}$ . Hence,  $a \in \text{RV}_\gamma$ , and

$$(5.2) \quad \frac{b(tx) - b(t)}{a(t)} \rightarrow \begin{cases} \frac{x^\gamma - 1}{\gamma} & \gamma \neq 0 \\ \log x & \gamma = 0 \end{cases}, \quad x > 0.$$

Inverting (5.2) gives

$$(5.3) \quad \frac{b^\leftarrow(a(t)x + b(t))}{t} \rightarrow \begin{cases} (1 + \gamma x)^{1/\gamma} & \gamma \neq 0 \\ e^x & \gamma = 0 \end{cases}, \quad x \in \mathbb{E}_\gamma.$$

Furthermore, if  $b^*$  is any function on  $(0, \infty)$  satisfying

$$(5.4) \quad (b^*(t) - b(t))/a(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty,$$

then (5.1), (5.2), and (5.3) hold with  $b$  replaced by  $b^*$ . A standard technique is to choose a smooth, strictly monotone  $b^*$  as is summarized next (cf. [21]).

**Lemma 5.1.** *There exists a function  $b^*$  satisfying (5.4) that is continuous and strictly monotone.*

**Proof.** Consider cases: If  $\gamma = 0$ , then  $b \in \Pi(a)$  and there exists [16, Proposition 0.16]  $\bar{b}$  continuous, strictly increasing such that  $(\bar{b}(t) - b(t))/a(t) \rightarrow 1$ . The choice  $b^*(x) = \bar{b}(e^{-1}x)$  satisfies (5.4). If  $\gamma > 0$ , then  $b \in \text{RV}_\gamma$ , and  $b(t)/a(t) \rightarrow \gamma^{-1}$  [7, Theorem B.2.2 (1)]. Consequently, [15, Proposition 2.6 (vii)] gives a continuous, strictly increasing function  $b^* \sim b$ . Writing

$$\frac{b^*(t) - b(t)}{a(t)} = \frac{b(t)}{a(t)} \left[ \frac{b^*(t)}{b(t)} - 1 \right]$$

shows that  $b^*$  satisfies (5.4). Finally, if  $\gamma < 0$ , then  $b(\infty) = \lim_{t \rightarrow \infty} b(t)$  exists finite,  $b(\infty) - b \in \text{RV}_\gamma$ , and  $(b(\infty) - b(t))/a(t) \rightarrow -\gamma^{-1}$ . Choose  $\bar{b}$  continuous, strictly decreasing, with  $\bar{b} \sim (b(\infty) - b)$ , and set  $b^* = b(\infty) - \bar{b}$ .  $\square$

It is easier to deal with  $b^*$  rather than  $b$  since  $b^{*\leftarrow}(b^*(x)) = b^*(b^{*\leftarrow}(x)) = x$  but  $b^*$  still standardizes  $Y$ . By (5.2),  $Y^* = b^{*\leftarrow}(Y)$  is in the standard domain of attraction when (5.1) holds:

$$tP[Y^* > ty] = tP \left[ \frac{Y - b^*(t)}{a(t)} > \frac{b^*(ty) - b^*(t)}{a(t)} \right] \rightarrow y^{-1}, \quad y > 0.$$



Furthermore if  $K(y, \cdot) = \mathbb{P}[X \in \cdot | Y = y]$ ,

$$(5.5) \quad K^*(y, \cdot) := K(b^*(y), \cdot)$$

is a version of the conditional distribution  $\mathbb{P}[X \in \cdot | Y^* = y]$ . This follows from

$$(5.6) \quad \mathbb{P}[X \in A, Y^* > y] = \int_{(y, \infty)} K(b^*(u), A) \mathbb{P}[Y^* \in du], \quad \text{measurable } A, y > 0.$$

To see this write,

$$\begin{aligned} \mathbb{P}[X \in A, Y^* > y] &= \mathbb{P}[X \in A, Y > b^*(y)] = \int_{(b^*(y), \infty)} K(u, A) \mathbb{P}[Y \in du] \\ &= \int_{(b^*(y), \infty)} K(b^*(b^{*\leftarrow}(u)), A) \mathbb{P}[Y \in du], \end{aligned}$$

using the fact that  $b^*(b^{*\leftarrow}(u)) = u$  for all  $u$ . Finish with a change variables to get (5.6).

We now show that the two approaches to the CEVM discussed at the beginning of Section 5, the direct approach and the standardization approach, are consistent.

**Proposition 5.1.** *Suppose  $Y$  has a distribution satisfying (5.1) and  $K^*$  is given by (5.5). Given normalization functions  $\alpha(t) > 0$  and  $\beta(t) \in \mathbb{R}$ , there exists a transition function  $\phi^* : (0, \infty) \times \mathcal{B}[-\infty, \infty] \mapsto [0, 1]$  such that, as  $t \rightarrow \infty$ ,*

$$(5.7) \quad K^*(tu_t, [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow \phi^*(u, [-\infty, x]) \quad \text{on } [-\infty, \infty]$$

whenever  $u_t \rightarrow u \in (0, \infty)$ , iff there is a transition function  $\phi : \mathbb{E}_\gamma \times \mathcal{B}[-\infty, \infty] \mapsto [0, 1]$  such that, as  $t \rightarrow \infty$ ,

$$(5.8) \quad K(a(t)u_t + b(t), [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow \phi(u, [-\infty, x]) \quad \text{on } [-\infty, \infty]$$

whenever  $u_t \rightarrow u \in \mathbb{E}_\gamma$ . If these convergences hold, then

- (i)  $\alpha, \beta \in \text{ERV}$ ;
- (ii)  $\phi^* = \kappa_{G^*}$ , a generalized tail kernel (4.3) with  $G^* = \phi^*(1, \cdot)$ ;
- (iii)  $\phi(u, A) = \kappa_G((1 + \gamma u)^{1/\gamma}, A)$ , where  $\kappa_G$  is a generalized tail kernel with  $G = \phi(0, \cdot)$ ; and
- (iv) the two transition functions are related by  $G = G^*$ .

**Proof.** Abbreviate  $a_t = a(t)$  and  $b_t = b(t)$ . The convergences (5.2) and (5.3) are locally uniform on  $(0, \infty)$  (see Section 2.1). Since  $b^*$  satisfies (5.4), it follows that

$$\frac{b^*(tu_t) - b_t}{a_t} \rightarrow \frac{u^\gamma - 1}{\gamma} \quad \text{whenever } u_t \rightarrow u \in (0, \infty),$$

and

$$\frac{b^{*\leftarrow}(a_t u_t + b_t)}{t} \rightarrow (1 + \gamma u)^{1/\gamma} \quad \text{whenever } u_t \rightarrow u \in \mathbb{E}_\gamma.$$

Assuming (5.7), for  $u_t \rightarrow u \in \mathbb{E}_\gamma$  we have

$$\begin{aligned} &K(a(t)u_t + b(t), [-\infty, \alpha(t)x + \beta(t)]) \\ &= K\left(b^*\left(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}\right), [-\infty, \alpha(t)x + \beta(t)]\right) \\ &= K^*\left(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}, [-\infty, \alpha(t)x + \beta(t)]\right) \\ &\Rightarrow \phi^*((1 + \gamma u)^{1/\gamma}, [-\infty, x]) =: \phi(u, [-\infty, x]) \end{aligned}$$

Conversely, if (5.8) holds, then for  $u_t \rightarrow u > 0$ ,

$$\begin{aligned} K^*(tu_t, [-\infty, \alpha(t)x + \beta(t)]) \\ &= K(a_t \cdot a_t^{-1}(b^*(tu_t) - b_t) + b_t, [-\infty, \alpha(t)x + \beta(t)]) \\ &\Rightarrow \phi(\gamma^{-1}(u^\gamma - 1), [-\infty, x]) =: \phi^*(u, [-\infty, x]) \end{aligned}$$

In either case,  $G := \phi(0, \cdot) = \phi^*(1, \cdot) =: G^*$ . Proposition 4.1 shows that  $\alpha$  and  $\beta$  are ERV and  $\phi^* = \kappa_{G^*}$ . Consequently,  $\phi(u, \cdot) = \kappa_G((1 + \gamma u)^{1/\gamma}, \cdot)$ .  $\square$

Therefore, by Proposition 4.1 (p. 9), if there exists a non-degenerate distribution  $G$  on  $[-\infty, \infty)$  such that

$$(5.9) \quad K^*(t, [-\infty, \alpha(t)x + \beta(t)]) = K(b^*(t), [-\infty, \alpha(t)x + \beta(t)]) \Rightarrow G(x)$$

with  $\alpha, \beta \in \text{ERV}$ , then (5.8) holds.

How can we apply Proposition 5.1 starting from an assumption like (5.9) on the kernel  $K$  rather than  $K^*$ ? Because  $b^*(b^{*\leftarrow}(t)) = t$ , (5.9) can be written as

$$K(t, [-\infty, \alpha \circ b^{*\leftarrow}(t)x + \beta \circ b^{*\leftarrow}(t)]) \Rightarrow G(x) \quad \text{as } t \rightarrow y^*,$$

where  $y^*$  denotes the upper endpoint of the distribution of  $Y$ , written as  $y^* = \sup\{y : F_Y(y) < 1\}$ . Therefore, we require there to exist a non-degenerate distribution  $G$  and normalization functions  $\tilde{\alpha} > 0$  and  $\tilde{\beta}$  such that

$$(5.10) \quad K(t, [-\infty, \tilde{\alpha}(t)x + \tilde{\beta}(t)]) \Rightarrow G(x) \quad \text{as } t \rightarrow y^*,$$

and  $\alpha = \tilde{\alpha} \circ b^*$ ,  $\beta = \tilde{\beta} \circ b^* \in \text{ERV}$ .

**5.2. CEVM Properties.** The standardization approach given in the previous section yields a CEVM when  $Y$  belongs to a general domain of attraction.

**Theorem 5.1.** *Suppose  $(X, Y)$  is a random vector on  $\mathbb{R}^2$ , where  $F_Y \in D(G_\gamma)$  according to (5.1) and  $K(y, \cdot) = \mathbb{P}[X \in \cdot | Y = y]$  satisfies (5.10) for normalizing functions  $\tilde{\alpha} > 0$  and  $\tilde{\beta} \in \mathbb{R}$  and non-degenerate limit distribution  $G$  on  $[-\infty, \infty)$ . Let  $b^*$  be the function satisfying (5.4) given by Lemma 5.1 and put  $\alpha = \tilde{\alpha} \circ b^*$ ,  $\beta = \tilde{\beta} \circ b^*$ . Then, as  $t \rightarrow \infty$ ,*

$$(5.11) \quad t \mathbb{P} \left[ \left( \frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)} \right) \in \cdot \right] \xrightarrow{v} \mu(\cdot) \quad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma),$$

where  $\mu$  is a non-null Radon measure satisfying the conditional non-degeneracy conditions (2.10), iff  $\alpha, \beta \in \text{ERV}_{\rho, k}$ . In this case, the limit measure  $\mu$  is specified by

$$(5.12) \quad \mu([-\infty, x] \times (y, \infty]) = \int_0^{(1+\gamma y)^{-1/\gamma}} G(u^\rho x + \psi(u)) du, \quad x \in \mathbb{R}, y \in \mathbb{E}_\gamma,$$

with  $\psi$  as in (2.2). The expression (5.12) is continuous in  $x$  and  $y$  if  $(\rho, k) \neq (0, 0)$ .

**Proof.** First, observe that  $Y^* = b^{*\leftarrow}(Y) \in D(G_1^*)$ . Defining the transition function  $K^*(y, \cdot) = \mathbb{P}[X \in \cdot | Y^* = y]$  as in (5.5), our hypotheses imply (5.9). Therefore, if  $\alpha, \beta \in \text{ERV}_{\rho, k}$ , then by Theorem 4.1, we have

$$t \mathbb{P} \left[ \left( \frac{X - \beta(t)}{\alpha(t)}, \frac{Y^*}{t} \right) \in \cdot \right] \xrightarrow{v} \mu^*(\cdot) \quad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty]),$$

where  $\mu^*$  is  $\mu^*([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} G(u^\rho x + \psi(u)) du$ ,  $x \in \mathbb{R}$ ,  $y > 0$ , and is conditionally non-degenerate. Consequently, for  $x \in \mathbb{R}$  and  $y \in \mathbb{E}_\gamma$ ,

$$\begin{aligned} t \mathbb{P} \left[ \frac{X - \beta(t)}{\alpha(t)} \leq x, \frac{Y - b(t)}{a(t)} > y \right] &= t \mathbb{P} \left[ \frac{X - \beta(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > \frac{b^{*\leftarrow}(a(t)y + b(t))}{t} \right] \\ &= \int_0^{(1+\gamma y)^{-1/\gamma}} G(u^\rho x + \psi(u)) du = \mu([-\infty, x] \times (y, \infty]), \end{aligned}$$

and the marginal transformation of  $Y$  does not affect conditional non-degeneracy or continuity. Conversely, (5.11) implies that  $\alpha, \beta \in \text{ERV}$  [10, Proposition 1].  $\square$

Alternatively, instead of standardizing  $Y$ , we could equally have used the convergence (5.8), which holds under our assumptions by Propositions 4.1 and 5.1.

Recalling the forms of the limit measure given in Section 4.1, we can express the limit measure in (5.12) as

$$\mu([-\infty, x] \times (y, \infty]) = \begin{cases} \frac{1}{\rho|x + k\rho^{-1}|^{1/\rho}} \int_0^{|x+k\rho^{-1}|(1+\gamma y)^{-\rho/\gamma}} u^{(1-\rho)/\rho} G(u \operatorname{sgn}(x + k\rho^{-1}) - k\rho^{-1}) du & \rho \neq 0 \\ \frac{1}{|k|e^{x/k}} \int_{-\infty}^{x \operatorname{sgn}(k) - |k|\gamma^{-1} \log(1+\gamma y)} e^{u/|k|} G(u \operatorname{sgn}(k)) du & \rho = 0, k \neq 0 \\ (1 + \gamma y)^{-1/\gamma} G(x) & \rho = 0, k = 0 \end{cases}$$

where  $\operatorname{sgn}(v) = v/|v| \mathbf{1}_{\{v \neq 0\}}$ , and we read the measure as  $(1 + \gamma y)^{-1/\gamma} G(-k\rho^{-1})$  when  $x = -k\rho^{-1}$  for the case  $\rho \neq 0$ .

In Example 4.1 (p. 11), we presented a transition function satisfying (4.1) which did not lead to a CEVM when paired with  $Y \in D(G_1^*)$ . We now show that a non-degenerate CEVM may be obtained if  $Y$  belongs to a non-standardized domain of attraction.

**Example 5.1.** Consider  $Y \sim \text{Exp}(1)$ , and  $U \sim \text{Uniform}(0, 1)$ , independent of  $Y$ . Put  $X = Ue^Y$ . Note that  $Y \in D(G_0)$  with  $a(t) \equiv 1$ ,  $b(t) = \log t$ , since for  $y \in \mathbb{R}$ ,

$$t \mathbb{P}(Y > y + \log t) = te^{-y - \log t} = e^{-y}.$$

A version of the conditional distribution is given by

$$K(y, [0, x]) = \mathbb{P}[X \leq x | Y = y] = \mathbb{P}[U \leq xe^{-y}] = xe^{-y} \wedge 1.$$

Taking  $\tilde{\alpha}(t) = e^t$ , we saw in Example 4.1 that

$$K(t, \tilde{\alpha}(t)[0, x]) \Rightarrow x \wedge 1 = G(x),$$

although  $\tilde{\alpha}$  is not regularly varying. Since  $b$  is continuous and strictly monotone, set  $\alpha(t) = \tilde{\alpha}(b(t)) = t$ . Then

$$K^*(t, t[0, x]) = K(b(t), \tilde{\alpha}(b(t))[0, x]) \Rightarrow G(x),$$

and  $\alpha(t) \in \text{RV}_1$ . Hence,  $K^*(tu, \alpha(t)[0, x]) \Rightarrow xu^{-1} \wedge 1 = G(u^{-1}x)$ , and  $K^*(y, \cdot) = \mathbb{P}[X \in \cdot | e^Y = y]$ . On the other hand, note that for  $u \in \mathbb{R}$ ,

$$K(a(t)u + b(t), \alpha(t)[0, x]) = txe^{-u - \log t} \wedge 1 = xe^{-u} \wedge 1 = G((e^u)^{-1}[0, x]).$$

This illustrates the equivalence presented in Proposition 5.1 (p. 16). Now, for  $x > 0$ ,  $y > 0$ , the joint distribution is given by

$$\mathbb{P}[X \leq x, Y > y] = \int_{\log x \vee y}^{\infty} x e^{-2u} du + \int_y^{\log x} e^{-u} du \mathbf{1}_{\{y < \log x\}}.$$

Therefore, for  $x > 0$ ,  $y \in \mathbb{R}$ , and large  $t$ , we have

$$t \mathbb{P}[X \leq tx, Y > y + \log t] = \begin{cases} \frac{x e^{-2y}}{2} & \text{if } \log x \leq y \\ e^{-y} - \frac{1}{2x} & \text{if } \log x > y \end{cases} = \mu([0, x] \times (y, \infty]),$$

and  $(X, Y)$  follow a CEVM by Theorem 5.1.

## 6. CONCLUSIONS AND FUTURE DIRECTIONS

In many statistical contexts a conditional formulation such as (4.1) is convenient. An example is when we model  $X$  as an explicit function of  $Y$  or when we work with distributions that have continuous densities, in which case the natural choice of version of the conditional distribution is the absolutely continuous one. In such cases, Heffernan and Tawn [11] approach is natural and leads to a parsimonious extremal model which can account for varying degrees of asymptotic independence. Heffernan and Tawn propose a semiparametric model, where the limit distribution  $G$  is estimated nonparametrically, and the normalization functions  $\alpha$  and  $\beta$  belong to a parametric family. The extended regular variation of  $\alpha$  and  $\beta$  provides justification for the form of the parametric family. The formulas for the limit measure derived in our present paper show assuming conditional distributions leads to a simpler CEV model parametrized by the distribution  $G$  and the pair  $(\rho, k)$ , along with  $\gamma$ , the extreme value index of  $Y$ .

Fitting a bivariate CEV model has been considered in [6, 9]. These authors discuss statistics for detecting a CEV model and estimating the normalizing functions. However, open questions remain, such as the asymptotic distributions of estimators, and the appropriate method for nonparametric estimation of  $G$ .

A natural extension of the bivariate model is to higher-dimensional vectors and this was the original intention of Heffernan and Tawn, who apply their methodology to a five-dimensional air pollution dataset. It is not clear how to condition on more than one extreme variable; presumably there are connections between such a model and the usual multivariate domain of attraction model. Cases where asymptotic independence is present between some pairs of variables but not others requires careful treatment and how to proceed in high dimensions is not apparent.

## REFERENCES

- [1] P. Billingsley. *Convergence of Probability Measures*. Wiley, 1968.
- [2] P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 2nd edition, 1999.
- [3] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 1989.
- [4] B. Das. *The conditional extreme value model and related topics*. PhD thesis, Cornell University, 2009.
- [5] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.
- [6] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.
- [7] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Verlag, 2006.

- [8] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.
- [9] A.L. Fougères and P. Soulier. Estimation of conditional laws given an extreme component. *Extremes*, 15(1):1–34, 2012.
- [10] J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571, 2007.
- [11] J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):497–546, 2004.
- [12] K. Maulik and S. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2004.
- [13] S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336, 2002.
- [14] S.I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1999.
- [15] S.I. Resnick. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer Verlag, 2007.
- [16] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 2008. Reprint of the 1987 original.
- [17] S.I. Resnick. Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics: An International Journal of Probability and Stochastics Processes*, 80(2-3):269–298, 2008.
- [18] S.I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1), 2013. To appear.
- [19] S.I. Resnick and D. Zeber. Clustering of Markov chain exceedances. Technical report, Cornell University, 2013. <http://arxiv.org/abs/1210.2314>; Submitted *Bernoulli*.
- [20] J. Segers. Multivariate regular variation of heavy-tailed Markov chains. *Arxiv preprint math/0701411*, 2007.
- [21] Eugene Seneta. *Regularly Varying Functions*. Lecture Notes in Mathematics, Vol. 508. Springer-Verlag, Berlin, 1976.

SIDNEY I. RESNICK, SCHOOL OF ORIE, RHODES HALL 284, CORNELL UNIVERSITY, ITHACA, NY 14853  
*E-mail address:* `sir1@cornell.edu`

DAVID ZEBER, DEPARTMENT OF STATISTICAL SCIENCE, 301 MALOTT HALL, CORNELL UNIVERSITY, ITHACA NY 14853  
*E-mail address:* `dsz5@cornell.edu`